Stat 524

11-28-17

①

<u>Midterm results</u>

$n = 35$ $\bar{x} = 40.1$ out of 48 (83.5%)

median = 43 (89.6%)

1	2
1*	
2	0
2*	5 7 9
3	0
3*	6 7 7 9 9
4	0 0 0 1 3 3 3 4 4
4*	5 5 5 5 6 6 6 7 7 7 7 7 8 8 8

②

In logistic regression, you have the model:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \varepsilon$$

where $p = P(Y=1)$

What does β_1 represent?

β_1 is the amount by which the left-hand side increases if x_1 increases by 1 unit and the other x-values remain constant

③

Let p_0 is the predicted probability when x_1 takes on a specific value x_0

Let p_1 be the predicted prob. when x_1 takes on the value $x_0 + 1$

$$\ln\left(\frac{p_1}{1-p_1}\right) = \ln\left(\frac{p_0}{1-p_0}\right) + \beta_1$$

$$\beta_1 = \ln\left(\frac{p_1}{1-p_1}\right) - \ln\left(\frac{p_0}{1-p_0}\right)$$

$$\beta_1 = \ln\left(\underbrace{\frac{p_1/q_1}{p_0/q_0}}_{\text{odds ratio}}\right) \qquad \text{so} \quad e^{\beta_1} = \frac{p_1/q_1}{p_0/q_0}$$

④

		Predicted 1	Predicted 0	
True	1	a	b	a+b
	0	c	d	c+d
		a+c	b+d	n

$$\text{Sensitivity} = \frac{a}{a+b} = \text{Prob}(\text{Pred } 1 \mid \text{actual } 1)$$

$$\text{Specificity} = \frac{d}{c+d} = \text{Prob}(\text{Pred } 0 \mid \text{actual } 0)$$

c is the # of false positives
b is the # of false negatives

5

A detail that was omitted from the discussion on ridge regression:

H & K suggested using the value

$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}, \text{ where } p = k+1$$

k: # predictors

$\hat{\sigma}^2$ = MSE
$\hat{\beta}$ = vector of coefficients

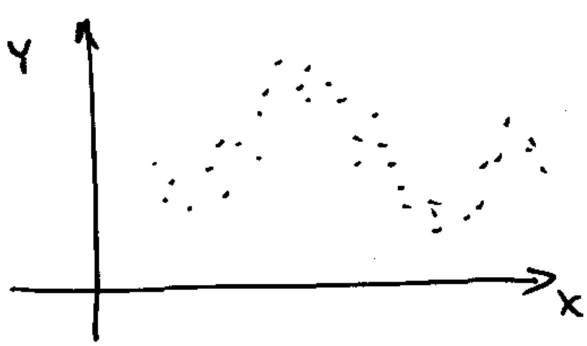
} from the OLS solution

R: lm.ridge

SAS: PROC REG (ridge = ...)

6

<u>Kernel regression</u> (model-free)



In OLS, $\hat{Y} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{H}Y = HY$

(7)

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & & \\ h_{n1} & \cdots & h_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\hat{y}_i = \sum_{j=1}^{n} h_{ij} y_j$$

Instead, define $\tilde{y}_i = \sum_{j=1}^{n} w_{ij} y_j$, $\sum_{j=1}^{n} w_{ij} = 1$

$$\tilde{Y} = W Y$$

(8)

How to construct the weights?

Use a <u>kernel function</u> $K(t)$

such that ① $K(t) \geq 0 \quad \forall t$

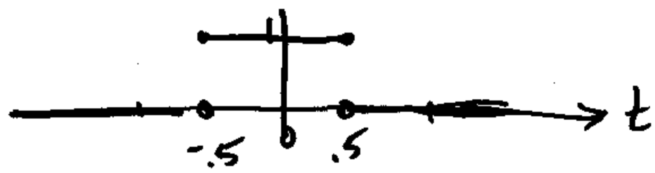
② $\int_{-\infty}^{\infty} K(t)\,dt = 1$

③ $K(-t) = K(t)$

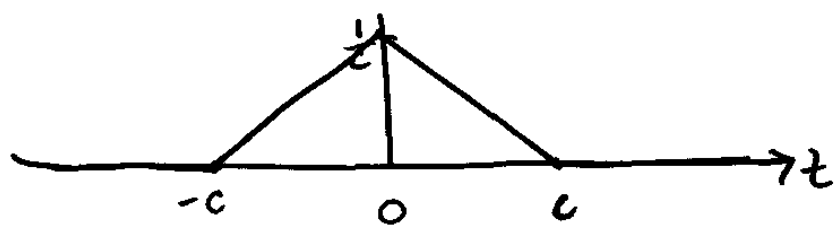
Examples of kernel functions:

(9)

Box: $K(t) = \begin{cases} 1 & |t| \le .5 \\ 0 & \text{otherwise} \end{cases}$



Triangle: $K(t) = \begin{cases} \left[1 - \frac{|t|}{c}\right]\frac{1}{c} & |t| \le c \\ 0 & \text{otherwise} \end{cases}$



(10)

Normal: $K(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{t^2}{\sigma^2}}$

Then define $w_{ij} = \dfrac{K\left(\frac{x_i - x_j}{b}\right)}{\sum_{k=1}^{n}\left(\frac{x_i - x_k}{b}\right)}$

b is called the "bandwidth"

Experiment with different kernels and bandwidths

(11)

LOESS Regression

(locally-weighted regression)

First, define a neighborhood by a <u>span</u>

for example, span = .3 $\Rightarrow$

use the closest 30% of the data to the point in question

Let x_i be a particular point, & let

$\Delta(x_i)$ = distance from x_i to the furthest point in its neighborhood

(12)

Let $$t_{ij} = \frac{|x_i - x_j|}{\Delta(x_i)}$$

Let $$w_{ij} = \begin{cases} (1 - t_{ij}^3)^3 & 0 \le t_{ij} < 1 \\ 0 & \text{otherwise} \end{cases}$$

Then use weighted L.S. to predict y_i.

Repeat this for each x value.